Google

DVAA Machine Costs
Nov 2022

Executive summary

- The problem
  ██

- Purpose of this Machine series
  - Understand DVAA machines costs at a deeper level. Look at comparables on how we could transition to a target level management of costs

- Summary of 2022 & 2023 baseline:
  - 2022: Compute & Storage (C&S) forecast is 1.0B, reflecting growth of +$38M (12%) vs 2021
  - 2023: C&S forecast is $1.2B (26% YoY growth)

- 2023 Plans:
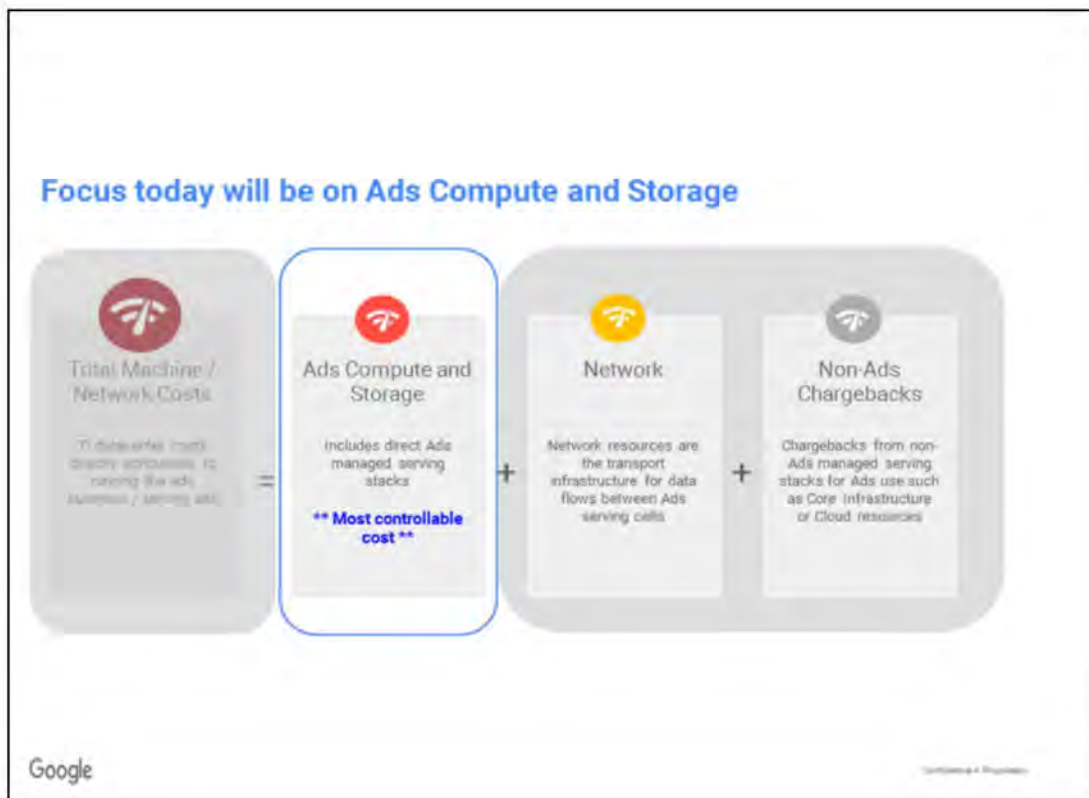 - Review opportunities to reduce expected machine costs in 2023

Google

GOOG-AT-MDL-008228529
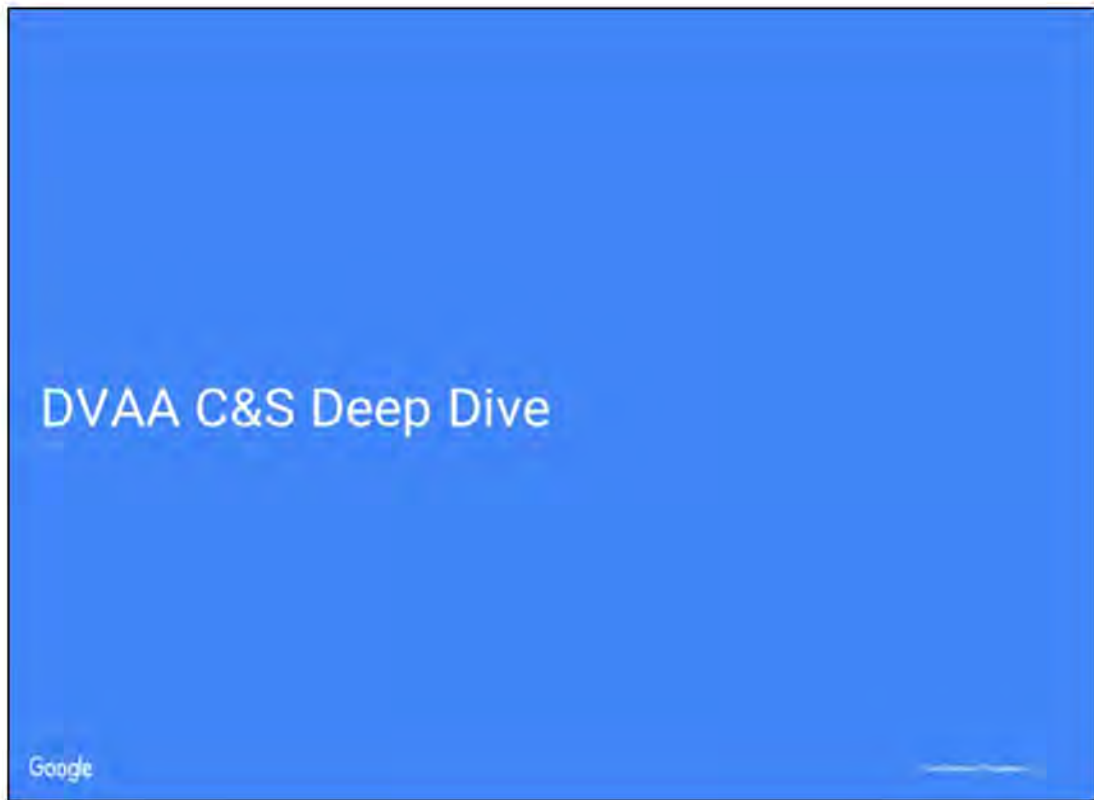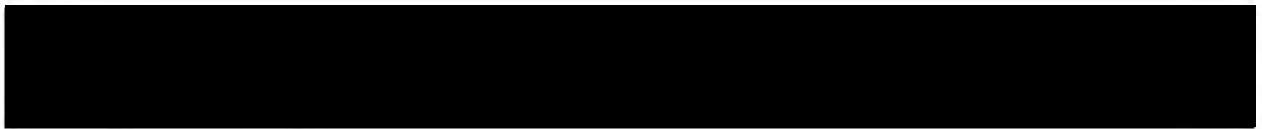
1. C&S: This includes cost streams nonML, ML, Logs, TI_Services, Lego, Dapa, Special Charges, and topsides.
2. nonML: tracks and plans for machine resources like GCU, RAM, Disk, SSD, spindles.
3. ML: Includes training and serving costs
4. ML training GCU: Same as "PA Planned nonML", but for anchors representing ML workloads: SmartAds*
5. ML training TPU: Costs from ML Fleet charged to Ads for training pods.
6. ML serving: Puffylite or Seastar accelerators used inside Ads for inference
7. Logs: Cost of storing and analyzing user activity logs
8. TI Services: include cloud services and services with service name in (annealing,midas,horizontal_monitoring_data,chronicle_on_blobstore,dremel, chronicle_on_blobstore_zonal,rdl,effingo,mindmeld,stackdriver,batch,forge,blo bstore,chronicle,frontend,scotty,st,google_service_control,timelock)
9. Lego: It represents a cost of holding MWs for future machine deliveries. When a machine is delivered this cost would be included into the machines' OPEX as a cost of power, and wouldn't be counted as Lego anymore.
10. Dapa: DAPA stands for Data Acquisition, Processing and Analysis. DAPA chargeback represents remuneration for Evenflow, Mesa, Napa, Rasta, F1 services. The majority of the cost & remuneration for these services occurs within the boundaries of the Ads PA, however some portions of these services are consumed by external PA groups & therefore the net cost for DIA chargeback is a negative cost.

GOOG-AT-MDL-008228533

11. Special Charges: Manual entries that are owned by TI, including adjustments, trueup, etc.
12. Topsides: Manual entries to the forecast owned by AdsRE
13. Core stream: Usually we just call it Core. It's one of the budget streams Ads has. It includes Core/search services.
14. Network: Network resources are the transport infrastructure for data flows between Ads serving cells. Machines need network to talk to each other, so PAs submit network resource orders along with compute & storage to the Google Technical Infrastructure org. Network is currently charged at the campus level (campus to campus flows) and managed at the Ads aggregate by the PARM team.
15. TI Tax: Tax charged TI Finance. (Link for more info)

16. Networking is charged back to PAs based on either Reservations (through B2/B4) or actual usage (Edge) – link
17. TI_Tax: is attributed to PAs based on a fixed percentage that the RE Finance team determines once all PA budgets are set for the subsequent year – link
18. 'Transformation Tax' is the delta we see when translating RIO PG costs into SAP PG costs – this is not unique to DVAA. For further information regarding that process I'd recommend reaching out to Connie Zeng

19. Plug is due to timing differences between real-time data found in go/ads-budget-model-2023 and snapshot used to feed into go/adspnl & go/projectslice.

20. Core(/Search) Chargeback are charged back to the PAs through Core chargeback - COST_TRANSPARENCY budget stream in Saber

21. Core Stream - P&L Attribution ratio is applied to the core chargeback from various anchors to different P&Ls

GOOG-AT-MDL-008228534

DVAA C&S Deep Dive

DTX-1375 Page 8 of 29

GOOG-AT-MDL-008228536

DTX-1375 Page 9 of 29

Display C&S Machine Costs by Cost Streams

	P&L Cost		Cost Difference		Total Growth		Run Rate		Incremental Growth	
	2022	2023	2022	2023	2022	2023	2022	2023	2022	2023

https://screenshot.googleplex.com/46AwpdygjeHh4uw - Display (11/1)

GOOG-AT-MDL-008228537

Id	Date	Text
1	11/04/2022 16:34:06	Hey Connie- these were pulled directly from go/ads-budget-model-2023 on 11/1 https://screenshot.googleplex.com/46AwpdygjeHh4uw
2	11/04/2022 16:57:32	I see. How was 2022 growth retrieved? Reason I am asking is that it may be a bit hard to obtain this information from one place (likely from two different dashboard with the current design) and prone to error. I can work with you to consistently populate this table going forward (I am working on a consolidated dashboard with this info)
1	11/04/2022 17:19:13	@cjalvarez@google.com @kieranknott@google.com How were the growth % numbers pulled? Can we check them together?
2	11/04/2022 17:19:13	2022 Growth was calculated in combination with a similar pull from go/ads-budget-model-2022 https://docs.google.com/spreadsheets/d/1aAeeZjA7kluyCaBDsUqe70AmC_RFCa07Nr5shV6qe5Q/edit#gid=1918901968

Google

GOOG-AT-MDL-008228538

Viewed by ██████, Display owned anchors account for nearly ████ of 2023 ████ growth

Google

GOOG-AT-MDL-008228539

Of [redacted] on Display P&L, Display owned anchors represent [redacted]

P&Gs USD Millions	Business Area	2022 Forecast	2023 Prelim	Y/Y $	2022 Y/Y %	2023 Y/Y %

Google

GOOG-AT-MDL-008228540

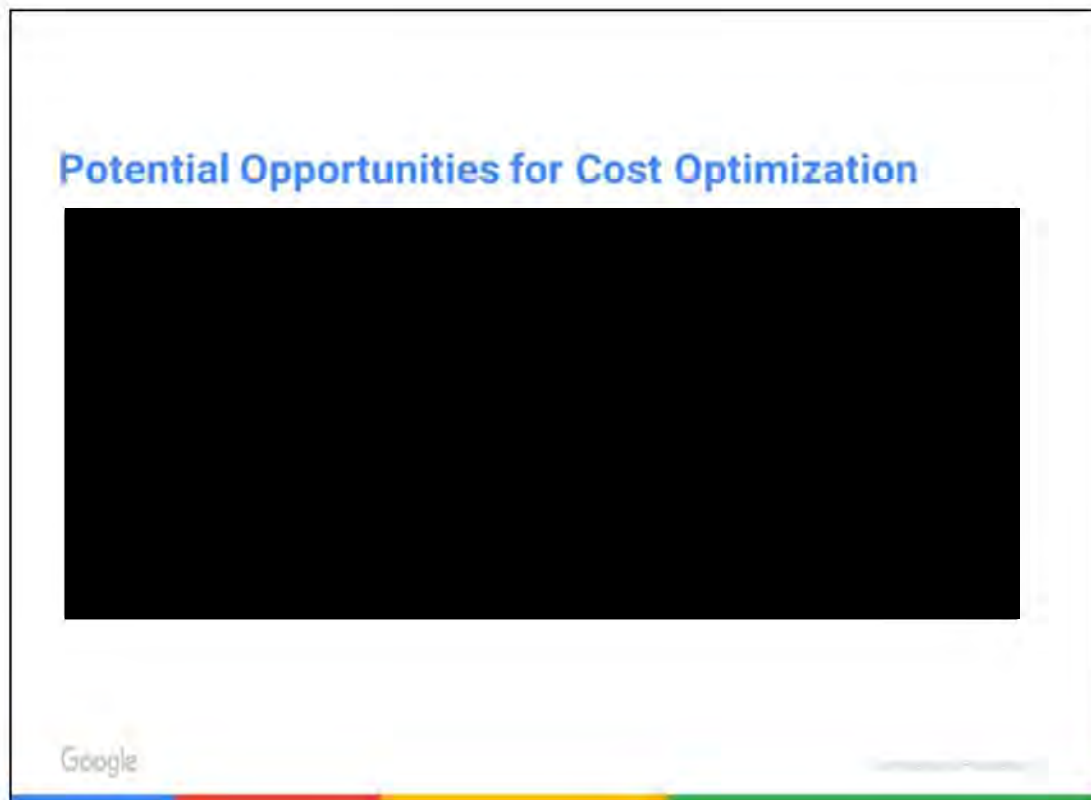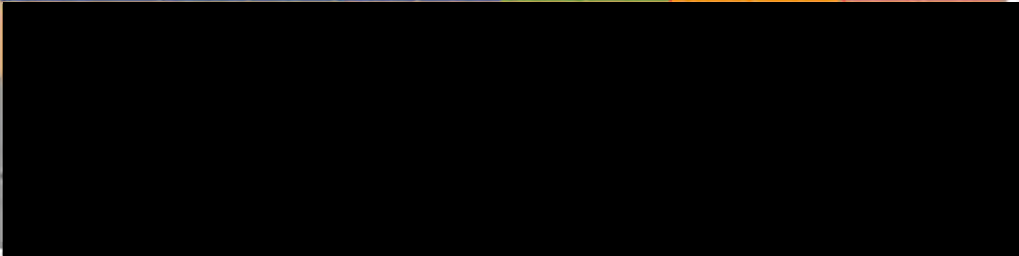| Id	Date	Text
1	11/22/2022 21:26:33	@qianyuzeng@google.com Hi Connie - as our teams look for savings opportunities to curb the growth on machines, is there a standard estimate we should use for converting SWE to $s? @vidyapedam@google.com FYI
3	11/22/2022 21:26:33	Let me ping you instead since this is a sensitive information

Google

GOOG-AT-MDL-008228541

Growth distributed among ▮▮▮▮ **while** ▮▮▮▮ **projected to remain flat**

Google

GOOG-AT-MDL-008228542

2023 Cost Savings

Google

Potential Opportunities for Cost Optimization

DTX-1375 Page 17 of 29

Google

GOOG-AT-MDL-008228544
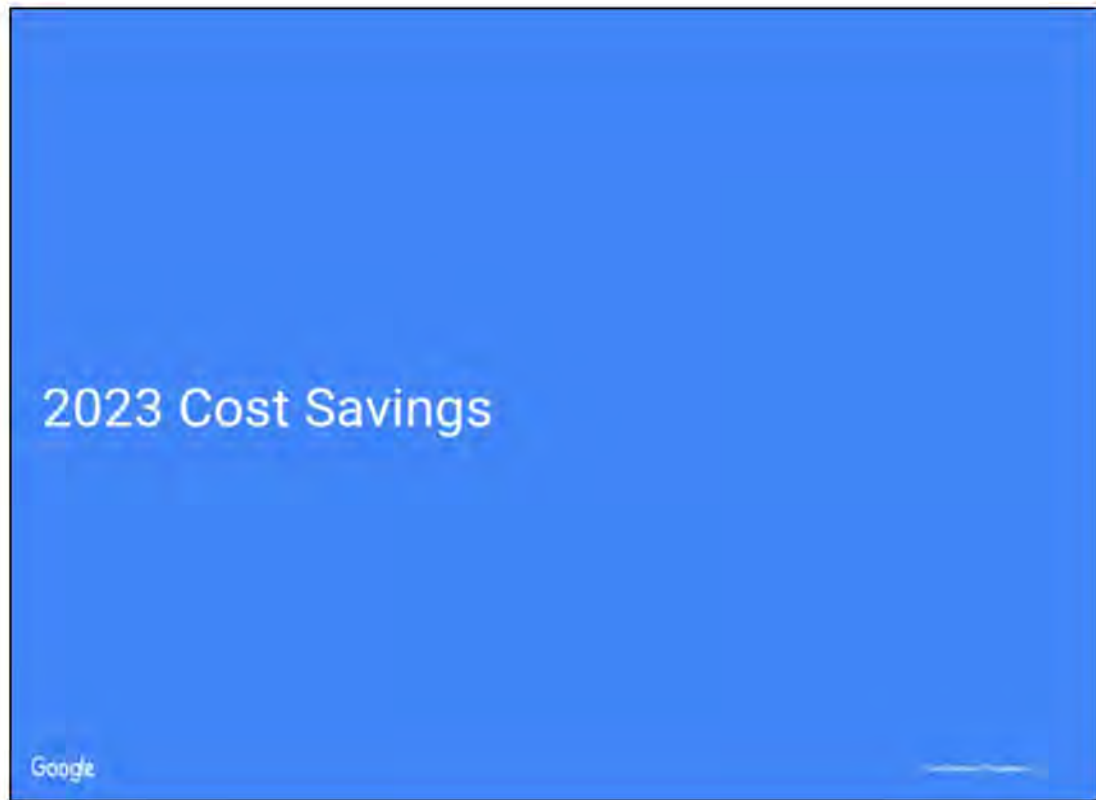
Machine Cost Reduction for DVAA P&L: Risk Analysis

| Display
$ in millions | FY 2022
Forecast | 2022 Y/Y
% | FY 2023
Prelim | 2023 Y/Y
% | FY 2023
[Revised] | 2023 Y/Y
% | FY 2023
[Revised] | 2023 Y/Y
% | FY 2023
[Revised] | 2023 Y/Y
% |
|---|---|---|---|---|---|---|---|---|---|---|

GOOG-AT-MDL-008228545

ML Cost Optimization to DVAA P&L

DTX-1375 Page 19 of 29

Google

Id	Date	Text
1	12/05/2022 17:40:09	@rajkusingh@google.com @vidyapedam@google.com will these actions really allow us to stop growth on GCU (and result in net reduction) - e.g. I thought the conclusion was that TidalStream migration is decoupled from this growth. Or is the statement that by mgirating to TidalStream, we'd be fully on TPU, so we wouldn't have to grow GCU anymore? While that's true, we'd presumably have to grow TPU further, so from a net budget perspective the impact is reduced. For opportunistic: do we know anything about likelihood of modeling teams to go that route? _Reassigned to Raj Singh_
1	12/05/2022 17:40:09	Tidal stream provides us a net reduction in cost. - TS TPU cost has been accounted for in the budget model but not the GCU returns. We should be able to return 914K GCUs ($5.6M) by Q2'23 After TS migration, we would still need some GCUs at a lower growth rate (eg. use case: Xborg GCUs would be required to cover input generation cost of Tidal stream models, support some models not migrated to TS) Can we modify the ask here to: "Provide support to accelerate TS migration for teams with layered training and TPU rescheduling due to slice fragmentation".
2	12/05/2022 19:33:47	@vidyapedam@google.com for all of these slides on cost stream can we add a top-level # to reinforce/remind what the 2023 net $ growth is + % growth at the top - this is to reinforce the baseline for which these cuts will be impacting - so that people don't have to remember state from a dense table in slides 9-11. Note that for this one the "Potential Business Risk" seems to put us in negative growth for ML - which we may not want to do in this particular domain. _Assigned to Vidya Priya Pedam_

Google

Log Cost Optimization to DVAA P&L

DTX-1375 Page 21 of 29

Google

GOOG-AT-MDL-008228548

Non ML Optimization to DVAA P&L

DTX-1375 Page 22 of 29

Google

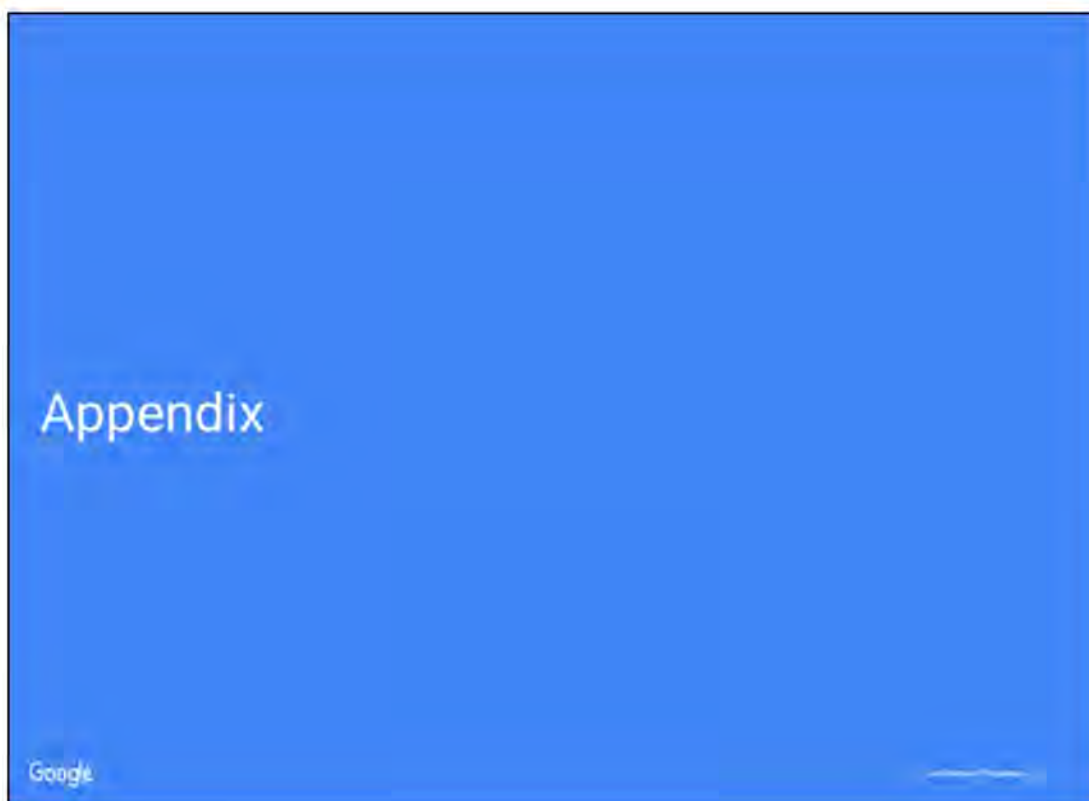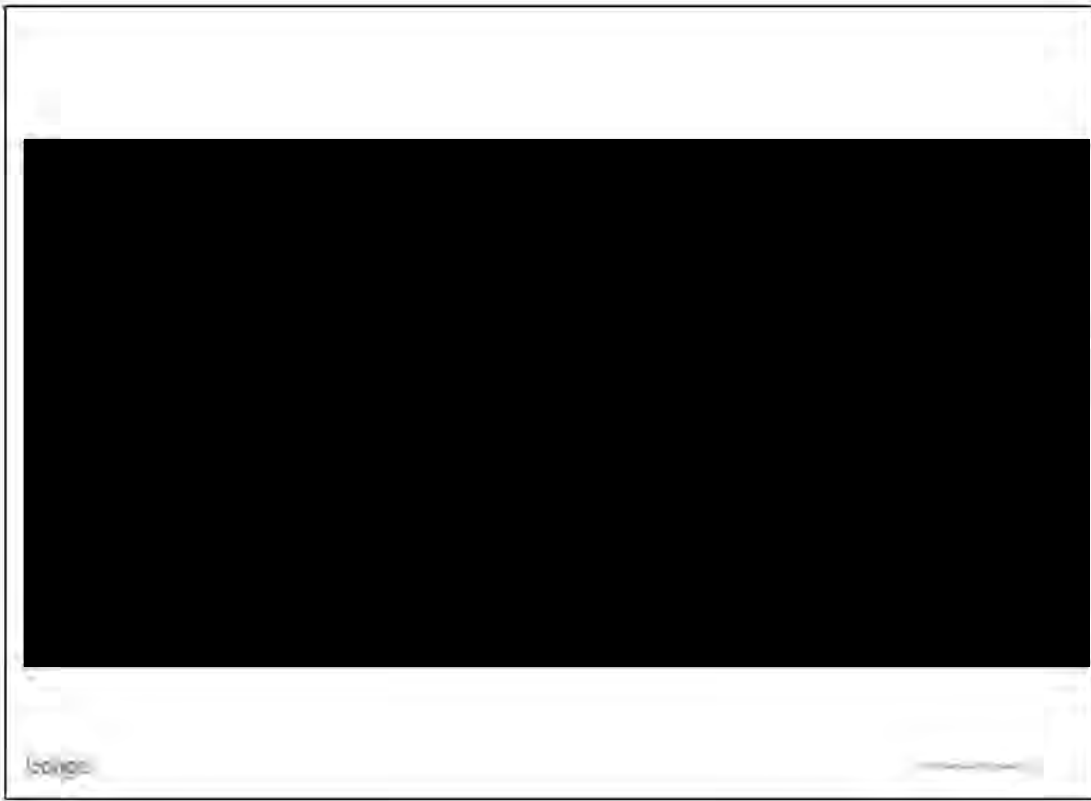| Id	Date	Text
3	12/05/2022 06:53:11	@edouard@google.com @vidyapedam@google.com @weihsinl@google.com @uminder@google.com Since this is the biggest "Potential Business Risk" bucket, let's move this and the Revenue SLO slide up to before ML and Log Cost - we want to make sure we get a steer on whether/how far we would want to go on Revenue SLO. _Assigned to Vidya Priya Pedam_
5	12/05/2022 19:51:04	@vidyapedam@google.com rephrased this for the audience: changing the tradeoff is the mechanism but we should surface what the key outcome is instead for this group. Can you update the Impact phrasing accordingly? _Reassigned to Vidya Priya Pedam_
2	12/05/2022 19:51:04	Hi Jason. I rephrased the impact. Does that capture the essence?
4	12/05/2022 19:52:02	Can we reframe "Asks" on this slide to Open Actions Required or something along those lines: this audience (Nima, Jai, Andy) are not necessarily the right people to make the decisions, so they aren't strictly Asks for the stakeholders.
1	12/05/2022 19:52:02	Thanks rephrased the title accordingly
6	12/05/2022 20:10:48	@vidyapedam@google.com can you show the relative breakouts of the impact you have here - since this is the biggest and highest-risk-inducing bucket, will be good to surface the details. _Reassigned to Vidya Priya Pedam_
3	12/05/2022 20:10:48	Added the breakdown

Google

Revenue SLO cuts

Details here.

present the option of giving up some revenue towards aligning our cost growth and revenue growth
Background is: we have the technical ability to throttle traffic and do so selectively. That is to say, we are able to predict the value of a query early, and use that information to decide whether to process that query or not.
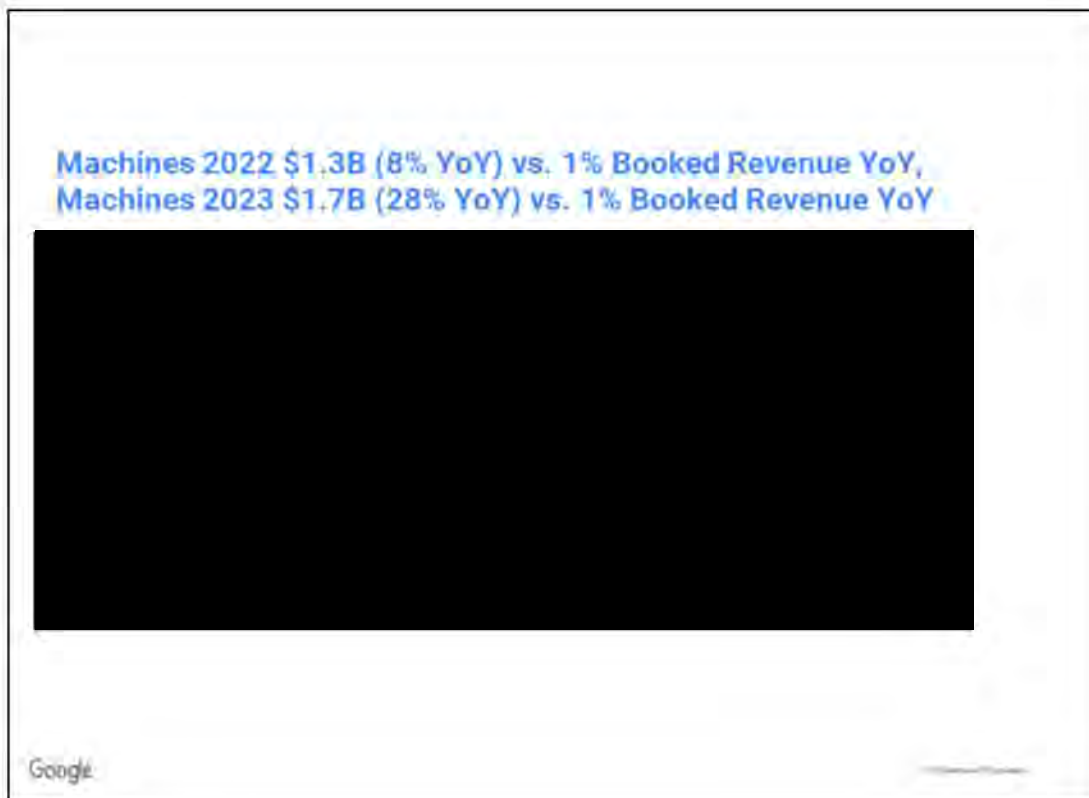By dropping low value queries, we can reduce cost growth faster than revenue growth.

In this table, we are presenting a few options:

DTX-1375 Page 25 of 29

GOOG-AT-MDL-008228552

DTX-1375 Page 26 of 29

GOOG-AT-MDL-008228553

Machines 2022 $1.3B (8% YoY) vs. 1% Booked Revenue YoY,
Machines 2023 $1.7B (28% YoY) vs. 1% Booked Revenue YoY

Google

1. C&S: This includes cost streams nonML, ML, Logs, TI_Services, Lego, Dapa, Special Charges, and topsides.
2. nonML: tracks and plans for machine resources like GCU, RAM, Disk, SSD, spindles.
3. ML: Includes training and serving costs
4. ML training GCU: Same as "PA Planned nonML", but for anchors representing ML workloads: SmartAds*
5. ML training TPU: Costs from ML Fleet charged to Ads for training pods.
6. ML serving: Puffylite or Seastar accelerators used inside Ads for inference
7. Logs: Cost of storing and analyzing user activity logs
8. TI Services: include cloud services and services with service name in (annealing,midas,horizontal_monitoring_data,chronicle_on_blobstore,dremel, chronicle_on_blobstore_zonal,rdl,effingo,mindmeld,stackdriver,batch,forge,blo bstore,chronicle,frontend,scotty,st,google_service_control,timelock)
9. Lego: It represents a cost of holding MWs for future machine deliveries. When a machine is delivered this cost would be included into the machines' OPEX as a cost of power, and wouldn't be counted as Lego anymore.
10. Dapa: DAPA stands for Data Acquisition, Processing and Analysis. DAPA chargeback represents remuneration for Evenflow, Mesa, Napa, Rasta, F1 services. The majority of the cost & remuneration for these services occurs within the boundaries of the Ads PA, however some portions of these services are consumed by external PA groups & therefore the net cost for DIA chargeback is a negative cost.

11. Special Charges: Manual entries that are owned by TI, including adjustments, trueup, etc.
12. Topsides: Manual entries to the forecast owned by AdsRE
13. Core stream: Usually we just call it Core. It's one of the budget streams Ads has. It includes Core/search services.
14. Network: Network resources are the transport infrastructure for data flows between Ads serving cells. Machines need network to talk to each other, so PAs submit network resource orders along with compute & storage to the Google Technical Infrastructure org. Network is currently charged at the campus level (campus to campus flows) and managed at the Ads aggregate by the PARM team.
15. TI Tax: Tax charged TI Finance. (Link for more info)

16. Plug is due to timing differences between real-time data found in go/ads-budget-model-2023 and snapshot used to feed into go/adspnl & go/projectslice.

17. Core(/Search) Chargeback are charged back to the PAs through Core chargeback - COST_TRANSPARENCY budget stream in Saber

18. Core Stream - P&L Attribution ratio is applied to the core chargeback from various anchors to different P&Ls

GOOG-AT-MDL-008228555

Id	Date	Text
4	11/03/2022 22:08:32	Hi Kieran, I added the description in the notes section. How did you query Core and Core chargeback? I wonder why the numbers are different.
4	11/03/2022 22:35:34	Hi Connie Core stream query can be found here: https://docs.google.com/spreadsheets/d/1aAeeZjA7kIuyCaBDsUqe70AmC_RFCa07Nr5shV6qe5Q/edit#gid=1322675551 in cell L7 Core charge back is provided here in Cell D8: https://docs.google.com/spreadsheets/d/1ohFczAzlY8d6vZjGdRGKkmLQMYrs1xLWM6pEAjTE3jo/edit#gid=751645743
5	11/03/2022 22:51:49	I see. Two different data sources and provided by two teams for the same thing, I imagine. The first one is managed by our team. You can reach out to aibekch@ or saraliu@ for more info. The second one is from P&L and managed by a different team? If we have both here, we maybe double counting it.
5	11/03/2022 23:12:03	What is the difference between the Core Chargeback vs. Core Stream? Core stream is now shown within go/ads-budget-model-2023; however, there is a core chargeback portion on our P&Ls as well (TI Deep Dive in go/adspnl: https://docs.google.com/spreadsheets/d/1ZqJBDkgOOb4wtgoU06U7RgpiCTo-xoBO2luvoBrKsYg/edit?resourcekey=0-O-4Fhlca9PKFVhk0-k_lxw#gid=2026725717) Core C&S is also found here: https://docs.google.com/spreadsheets/d/1zZUm1zcjWMhtKB7I4-L62ChT0V-uQU6aqo6PfyHZeTY/edit#gid=1817104967
6	11/03/2022 23:21:56	@saraliu@google.com @aibekch@google.com can you please take a look at Kieran's questions about Core? Thanks.
3 Google	11/04/2022 15:54:29	Hey @qianyuzeng@google.com Can you provide a brief description for each line item in the Notes section below? Additionally, could you provide an explanation of the difference between (1) Core Stream vs. Core Chargeback, (2) ti_tax% vs. ti_services? Confidential + Proprietary Thanks

GOOG-AT-MDL-008228556